DOCUMENT RESUME

ED 291 251 FL 017 179

AUTHOR Roberts, A. Hood
TITLE The System of Communication in the Language Sciences: Present and Future. LINCS Project Document Series. LINCS #19-69P.
INSTITUTION Center for Applied Linguistics, Washington, D.C.
SPONS AGENCY National Science Foundation, Washington, D.C.
PUB DATE Oct 69
GRANT GN-771
NOTE 21p.; Paper presented at a Johns Hopkins University Conference on Communication among Scientists and Technologists (October 28-30, 1969).
PUB TYPE Information Analyses (070) -- Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Clearinghouses; *Communication Problems; Information Needs; *Information Networks; Information Science; Information Transfer; Interprofessional Relationship; *Linguistics; Program Design; Trend Analysis
IDENTIFIERS *Language Information Network Clearinghouse System

ABSTRACT

A discussion of communication within the field of language sciences looks at the existing system of communication, noting some of the major problems and outlining some special and strategic problems facing those who design future information systems in this area. The report begins by describing the language sciences community and its communication patterns, and then focuses on existing information sources, developmental activities, and some problems of communication system design. The development and early results of the Language Information Network and Clearinghouse System, in existence for only two years at the time of writing, are discussed. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED291251

CENTER FOR APPLIED LINGUISTICS

LANGUAGE INFORMATION NETWORK AND CLEARINGHOUSE SYSTEM (LINCS)

THE SYSTEM OF COMMUNICATION IN THE LANGUAGE SCIENCES: PRESENT AND FUTURE

By A. Hood Roberts

In Collaboration with

LINCS Project Personnel and Consultants

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

GRT

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

For Presentation to Johns Hopkins University Conference on Communication among Scientists and Technologists, October 28-30, 1969

Not for General Circulation

LINCS PROJECT DOCUMENT SERIES / NATIONAL SCIENCE FOUNDATION GRANT

LINCS #19-69 P October 1969 NSF GN-771

CENTER FOR APPLIED LINGUISTICS, 1717 MASSACHUSETTS AVENUE, N.W., WASHINGTON, D.C. 20036

FL017179

BEST COPY AVAILABLE

THE SYSTEM OF COMMUNICATION IN THE LANGUAGE SCIENCES: PRESENT AND FUTURE

By A. Hood Roberts

In Collaboration with
LINCS Project Personnel and Consultants

For Presentation to Johns Hopkins University Conference on Communication among Scientists and Technologists, October 28-30, 1969

CONTENTS

1. Introduction 1
2. The Language Sciences Community 1
3. Existing Information Resources 7
4. Developmental Activities: The LINCS Project 12
5. Some Problems of System Design 14

1. Introduction

The intent of this paper is to highlight the distinctive aspects of communication in the language sciences, to describe briefly the existing system of communication with special attention to some of the major problems, and to sketch some of the strategic and special problems facing the designer of a future information system in this area. It is appropriate to note, at the outset, our present point in a progression of exploratory and developmental activities at the Center for Applied Linguistics (CAL): the Language Information Network and Clearinghouse System (LINCS) project -- the chief locus of these activities -- has been in existence only since July 1967. Although many survey findings are now becoming available, the overall system-design effort has not yet proceeded beyond a general consideration of preliminaries. It is, however, reasonable to claim that most of the essential user needs studies have been completed, with the exception of some refinements covering active scientists, certain language specialists, and various practitioners including translators. The base line attained is certainly adequate for marketing studies of specific products and services, which are currently under way.

This presentation begins with an operational outline of the language sciences community and its communication patterns. Thereafter, the discussion focuses on existing information resources, current developmental activities, and some problems of system design that may be of special interest to this audience.

2. The Language Sciences Community

The construction of a working definition of the scope of the language sciences would seem to be a logical first step in a project of the nature of LINCS. It has been our feeling, however, that such a definition should be functional rather than theoretical; an attempt on our part to define at this point in time the theoretical boundaries of the language sciences would not only be premature, but would probably diminish, rather than enhance, the usefulness of whatever system we may ultimately develop. For present purposes we consider the language sciences to embrace all fields of study which pertain to the systematic examination of human language and communication. This is admittedly very broad, but intentionally so: our purpose is to be as unrestrictive at the outset as possible.

The multitude of subject areas covered by this definition have been conceptualized as three concentric circles. At the center of this pattern lies linguistics, which is concerned directly with the study

of the sounds, structures, and vocabulary of all languages, as well as their dialects, their genetic and social interrelations, and so forth. Language learning and teaching would also be considered part of this core. The accumulation and analysis of social, anthropological, and psychological information about the speakers of languages lie on the outer fringe of the core and lead into the next area -- of cross-disciplinary specialties. These include the psychology and sociology of language, acoustics, certain computer applications of language, stylistics, and the fields concerned with pathologies of human communication, and speech behavior. This second concentric ring is distinguished from the core by having language as its basic subject, but bringing to the study of language a consideration of other disciplines. The outermost circle includes those fields which are oriented towards language as a tool: these include symbolic logic, information science, information theory, translation, graphics, experimental psychology, psychiatry and mass communication, among many others. However, this comprehensive conceptualization of the language sciences does not necessarily imply that a future discipline-oriented information system will give equal weight to all subject categories. It is very likely that peripheral topics will be covered by cooperative arrangements for the exchange of information with other disciplines.

In terms of our functional definition, the language sciences community is defined as being composed of those persons whose professional activities and interests bear on the study of human language and communication in any form. From studies of the quantitative and the dynamic aspects of this community a number of interesting and distinctive traits have emerged. The first of these is perhaps the sheer size of the community. An early estimation placed the figure for a total potential United States audience for a LINCS at about 100,000. More recent estimates -- still conservative -- have raised that figure to about 200,000 for the United States. Of these 200,000 individuals, approximately 6,000 are specialists in linguistics, the core discipline. Of the remainder, over 100,000 are teachers of English or foreign languages.

The large number of language teachers brings out a fact that should be noted in consideration of the schematization of the language sciences described earlier. The placement of a particular specialty in the tripartite conceptualization of the language sciences has no necessary bearing on the relative importance of this specialty as a component of the future clientele, from a marketing point of view. In addition to its large size, the audience for an information system in the language sciences is highly diversified in subject concerns and professional and scientific activities. This obviously could have been foreseen from the breadth of our description of the scope of the language sciences. The difficulties engendered by this

diversification, however, could not have been avoided by a narrower definition of the field: information generated in one of the language sciences may have, either in its original form or in some permutation, a very high transfer value for several other fields.

Not only is the potential audience highly diversified in its activities and subject-matter interests; its pattern of membership is highly fragmented in terms of the variety of professional and scientific organizations -- in most cases with a relatively limited mandate. Such fragmentation does not correspond only to differences in subject matter; within any given subject area the same phenomenon may be found to a high degree. It may best be seen from a study of the professional organizations and societies which have relevance to the language sciences. One of our project's current activities is aimed at the compilation of an inventory of such societies with a view towards future collaboration with them. As a first step, a list of about 70 societies in the United States has been compiled. Of these, about half seem to have some area of the language sciences as their primary emphasis. Current and future studies will undoubtedly expand the list; our best projection is that the ultimate list will cover about 200 societies of national membership.

We have found that the amount of overlap among the societies with interests germane to the language sciences is much lower than might logically have been expected. (It was this fact which necessitated our reestimation upwards of the size of the total community.) For example, we expected to find a heavy overlap between the membership of the Modern Language Association of America (MLA) and the American Association of Teachers of French (AATF), and between the MLA and the American Association of Teachers of Spanish and Portuguese (AATSP). Instead, it has been discovered that only about 10% of the members of each of these two other societies also belong to the MLA. Furthermore, the figures gathered to date do not take into account three more or less "fugitive" segments of the community: students, translators, and other application-oriented persons, and professionals who do not join societies.

There does not exist at this time any umbrella organization for the language sciences in general. This being true, no language science society functions as a guild, and there is therefore no great pressure to join any particular society. Estimation of the size of the "invisible" segments of the language sciences community is therefore extremely difficult. Investigations of translators and other applied workers are in the planning stage; we hope in the near future to have at least preliminary data on these sub-populations.

There are 4,000-6,000 languages in the world and any of these, or any dialect of these, may be the subject of linguistic studies. In addition, the aspect studied may correspond to any of the various

subspecialties in linguistics, and may be further qualified by the theoretical orientation of the scholar. The number of possible combinations of these factors is obviously huge. (The specialty most frequently encountered among linguists is the study of the structure of some particular language or language group.) This has bearing not only on the variegation of information of interest to the language scientist; it affects also the processing techniques and media utilizable in providing that information. I will have more to say about this when I come to the problems facing the system designer.

Despite diversification on the points already mentioned, some points of homogeneity can be noted among language scientists, at least in the core group of linguists. (These factors have emerged from an analysis of the linguistics section of the National Science Foundation's National Register of Scientific and Technical Personnel for 1968; this section has been administered by the Center for Applied Linguistics since 1964.) These data give only a partial view of the field, since they are limited to American-born or resident scientific linguists. A first observation on the responding population is that despite wide heterogeneity in subject specialties, there is considerable similarity in the matter of professional activity. The large majority -- over 70% of the respondents -- were university-based and divided their time fairly evenly between teaching and research, with teaching running somewhat ahead. A surprisingly large proportion of the respondents were involved in management: 11% listed this as their primary work activity and 11% as their second most important work activity. The population of linguists seems to be quite spread out geographically: the respondents were widely distributed throughout the United States. In addition, over 10% of the respondents, although American-born, were residents in foreign countries. It seems unlikely that this statistic would be duplicated in any other field.

One finding we expected when we began the analysis of the National Register data was not corroborated. Those familiar with the field of linguistics have been struck for some time by the extent to which differences in theoretical orientation seem associated with differences in age; we seem to be confronted with a series of "generation gaps". Although this phenomenon has, to some extent, been observed in the preliminary behavior studies we have conducted to date, it was not at all borne out in our analysis of the National Register material. The conclusion seems to be that this fact is more or less limited to highly active, creative individuals, and does not represent a general state of affairs within the field. The only feature which seems to distinguish younger linguists in general from their older colleagues is a growing eclecticism.

In addition to the foregoing efforts, we have initiated studies of the dynamics of information generation, processing and transmission, with a particular emphasis on informal practices. Like our other studies, these have concentrated to date on those segments of the community most directly concerned with core subject specialties, but further investigations, covering additional components of the total population, are in the planning stage.

Preliminary data are now available from a survey of information practices and needs of members of the Linguistic Society of America (LSA), which confirm the impression left by the National Register data of a university-based, heavily teaching- and research-oriented group with widely diversified interests. These people appear to spend relatively more time in teaching than do persons in the physical or social sciences: three-quarters of the foreign language and English specialists in the study ranked teaching first in time consumption. The subjects of this study relied on a wide range of media both to locate and obtain information. The media most widely used in obtaining information were books, journals and discussions with colleagues. In locating information, however, there was a much more widespread reliance on a great variety of media: citations, scanning the periodical literature, critical reviews, bibliographies, abstracts, etc. The respondents in this study listed teaching and research with almost identical frequency as the activities making the greatest demands on them in the gathering and use of information related to language. The subject areas considered very important sources of required information were quite varied. As might have been expected, the most frequently cited were linguistics, scholarship in a particular language or language family, English, and language teaching methodology. Some 50 languages were mentioned in this connection. A variety of other areas were also cited, however, ranging all the way to medicine, mathematics, and computer programming. It must be borne in mind, in this connection, that we are now discussing people interested in the topical core of our field: certainly the incidence of interest in such wide-ranging fields may be expected to increase in segments of the language sciences community less directly concerned with linguistics.

An interesting datum highlights nicely this heterogeneity of interests and needs of language scientists. Subjects of the study were asked to name the journals they would like to see covered in a hypothetical current awareness service: the 349 respondents named a total of 329 different journals.

A case study of communication practices in the language sciences in the Washington, D.C., area has recently been completed; its results are being worked into a format appropriate for informal distribution. Using both a questionnaire survey and 70 personal interviews, this

study examined the self-identification, training, interests, and information needs of language professionals in and around Washington. Both parts of this study seem to reinforce the picture that has begun to emerge from other researches. The subjects were so heterogeneous that, for interpretation of data, they had to be divided into several groups, according to subject area (linguistics, specialization in languages, specialties in other fields, association with common or exotic languages), and, for some purposes, work activity. Respondents differed in age, degree level, self-identification, type of institution at which they were employed, interest in scientific or humanistic aspects of language (linguists were scientifically oriented; foreign langage teachers were more interested in literature), and so on.

Despite very distinctive differences in discipline orientation, the subjects seemed to hold some characteristics in common. About two-thirds of them, in all categories, relied heavily on primary publications for professional information; slightly less on colleagues. Three-quarters of all the subjects used various secondary sources regularly; as a secondary source, colleagues ranked very high. One reason for such active use of colleagues as an information source seemed to be that many of the subjects felt published sources to be inadequate. The precise role of these various sources shifted according to the group differences mentioned previously, but overall, there was considerable complaint that needed data and important documents were inaccessible.

Informal communication seems to crop up as a particularly important activity among the interview subjects (this might be partially attributable to the interview method). Such activity has been studied in greater detail in another interview study, preliminary in nature, whose objective was the formulation of a plan for a large-scale study of informal information exchange among active linguists. The pilot study was a series of interviews with 13 eminent and highly productive linguists on the East Coast. The subjects were questioned at length about both formal and informal channels of information transfer. Despite the small number of subjects, heterogeneity was, once again, a key characteristic: nine of the interviewees clustered into three basic patterns, based on subject-matter interests, and information use and exchange, while each of the remaining four had some characteristic so distinctive as to make him unclassifiable. Interestingly enough, the generational differences in approach, training, and interest that I referred to earlier -- those not borne out by our analysis of National Register data -- emerged here. Specifically, younger researchers were mainly oriented to transformation theory: middle-aged and older research workers were more oriented to "general" or "Bloomfieldian" linguistics or to philological scholarship. There emerged in all the behavioral studies a surprising number of problem areas that attract only about a half dozen scientific workers. The

findings of this limited study will be used to plan a more exhaustive investigation of the communications practices of highly active linguists; from that point, we will go on to study the behavior of persons active in other of the language sciences.

Finally, in our studies to date of the language sciences community, we have attempted two "unobtrusive" studies of information transfer in linguistics. One of these examined the volume of material generated by a number of research projects, the methods of dissemination used, and time factors involved. I will mention a few of the data found in this study in a moment, when I come to the description of the existing information resources situation. The other study, which is still in progress, is of citation patterns in twelve "core" linguistics journals. The study involves about 3,000 citations, and is utilizing clustering techniques to investigate patterns among the sources of citations, types of literature cited, the authors of cited articles, and chronology trends. From this we hope to shed more light on the internal structure of communication within the core field of linguistics, as well as its place within the wider context of disciplines concerned with language.

3. Existing Information Resources

In their more general aspects, the formal channels for the transmission of information in the language sciences resemble those in most other scientific disciplines. The same arrays of primary, secondary, review, and institutional publications are to be found here as elsewhere, and they may generally be said to have the same fundamental virtues and defects. The study of research reporting that I mentioned a few moments ago suggests that outlets used most frequently are journal articles, conference papers, and technical reports. Slightly over half of the items included in that study were covered in widely available abstract journals, bibliographies, and indexes. The same study indicates that about two to four and a half years elapse between initiation of a research report and its publication in a journal or in conference proceedings. Approximately one and a half years after this, a little over 50% of the items have been covered in secondary publications. Although the proportion of secondary coverage may be relatively low, this general picture does not seem particularly different from circumstances observable in other disciplines.

In certain respects, on the other hand, information resources in our field are quite distinctive. First, the literature appears in a far greater variety of languages than may be found elsewhere,

about 70. Even for linguists, this poses a problem. Second, the distribution of the literature among these languages is more even than that found in other areas -- particularly the "hard" sciences, where English has such clear hegemony. Although English accounts for a larger portion of the literature than any other single language, its predominance is not so marked as elsewhere; about six languages account together for the majority of the world's output. Furthermore, languages that in other fields account for very little of the significant formal literature, in linguistics have unusual importance (e.g. Czech). Third, the secondary and tertiary (review) publication systems are not so well developed in the language sciences as in other subject areas.

A preliminary survey by the LINCS project put the number of periodicals relevant to the language sciences at something over 2,000, with about 250 of these being "core" ("very high yield") linguistics journals, and another 100 "high yield". By all odds, these figures are highly conservative. A new serials inventory, just now beginning, will provide a more accurate view of the field.

A word on problems associated with the "peripheral" literature might be in order here. It is particularly true in an area as interdisciplinary as the language sciences that the quality of being peripheral (i.e. published in publications not located near the center of the concentric configuration of disciplines I referred to at the outset of this talk) has to do only with an item's visibility -- not its relevance. In the study of members of the Linguistic Society of America which I mentioned earlier you will recall that the subjects nominated journals for a current awareness service. Of the 35 most frequently nominated, many were not "core" linguistics journals.

As I have already stated, the secondary publication system in the language sciences is not particularly well developed. Worldwide, only about 40 secondary publications process a significant volume of pertinent material. There are two major annual bibliographies, containing about 12,000 items each -- one with a two and a half year publication lag (i.e. it actually appears two and one half years after the date on its cover). Overlap in coverage between them is 30-40%, so that there is considerable duplication of effort. There is no central abstracting service to cover all of linguistics, let alone all the language sciences. Language and Language Behavior Abstracts[1] covers articles which approach language from an interdisciplinary point of view but does not cover those of interest only

---

[1] Harlan L. Lane et al., eds., LLBA: Language and Language Behavior Abstracts (Ann Arbor: The University of Michigan Center for Research on Language and Language Behavior (CRLLB), 1967—, quarterly).

to scholars working in a single discipline. In the behavioral studies described earlier, subjects were asked to evaluate various tools for the location of information. Reactions to abstracts were notably mixed; incidence of their use was not as high as might have been expected, and a number of respondents complained of difficulty in locating abstracts, presumably because of the absence of any central service. In point of fact, linguists are not really in a position to evaluate abstracts as a tool, given their general dearth and the restricted coverage of the abstract publications that do exist. Bibliographies of various kinds -- special and general, annotated, indexed, and not -- have until now constituted the major secondary instrument available to the language scientist. None of these attempts to cover the entire literature, and, as has been said, even the most important of them suffer from severe time lags. There is, however, a profusion of them: a "bibliography of bibliographies" lists more than 2,000 in the Soviet Union alone. Duplication, needless to say, is very high.

Relatively speaking, there is very little tertiary (review) literature to consider. The Educational Resources Information Center (ERIC) Clearinghouse for Linguistics, located at CAL, has produced some state-of-the-art reviews. Some specialized topics are covered in the review publications of other disciplines; otherwise there is little to mention.

Information centers, on the other hand, have been springing up fairly quickly. The LINCS project has undertaken a worldwide descriptive survey of such centers; according to our findings, there are now about 100, varying widely in size, affiliation, and function. In this connection, we might note that CAL itself was established, in large part, to provide the services of an information center, and has, in addition to the foundation of the LINCS project, developed a variety of services. In addition to housing the ERIC Clearinghouse, already mentioned, it offers a newsletter, The Linguistic Reporter[2], and a current awareness service called Language Research in Progress. It initiates many special publications, maintains an extensive library, and responds to thousands of queries yearly. It has collaborated with the Permanent International Committee of Linguists to improve the international Linguistic Bibliography[3], and also with

---

[2] The Linguistic Reporter (Washington, D.C.: Center for Applied Linguistics, 1959—, six numbers a year).

[3] Permanent International Committee of Linguists, Linguistic Bibliography for the Year 19— and supplement for previous years (Utrecht: Spectrum, 1949—).

the compilers of the annual bibliography of the Modern Language Association of America[4]. In a more general way, it has worked to improve international cooperation and coordination, sharing of resources, and modernization of techniques of literature control in the language sciences.

The University of Michigan's Center, already mentioned, is a cooperative venture that utilizes a small network of specialized information centers. In addition to its abstract journal and announced review series, it provides a (limited) reprint service and a directory of journals; it has plans to publish its thesaurus and operate a retrieval service.

In England, the Centre for Information on Language Teaching (CILT), a government-supported foundation, is concerned with the collection, coordination, and dissemination of information on all aspects of modern languages and their teaching. In conjunction with the English-Teaching Information Centre of the British Council it covers the book literature as well as about 300 periodicals, and publishes Language-Teaching Abstracts[5] and A Language-Teaching Bibliography[6], in addition to maintaining a register of current research in Great Britain, which is modeled on CAL's Language Research in Progress System.

A final illustration of the kinds of information centers developing in the language sciences is the very competent Information Center for Hearing, Speech, and Disorders of Human Communication of the Johns Hopkins University. It demonstrates quite strikingly the interdisciplinary nature of the language sciences, drawing its input from a broad range of subject specialties in a variety of media, and serving a number of divergent interests in the biomedical community. It provides material in current awareness services, specialized bibliographies, reviews, and state-of-the-art reports. The Center is one of several in the Neurological Information Network of the National Institute of Neurological Diseases and Stroke.

---

4 MLA International Bibliography of Books and Articles on the Modern Languages and Literatures for 19— (New York: New York University Press, 1956—).

5 English-Teaching Information Centre of the British Council and the Centre for Information on Language Teaching, comps., Language-Teaching Abstracts (Cambridge and New York: Cambridge University Press, 1968—, quarterly).

6 Centre for Information on Language Teaching and the English-Teaching Information Centre of the British Council, comps. and eds., A Language-Teaching Bibliography (Cambridge: At the University Press, 1968).

Approximately 30 of the 100 information centers identified for the language sciences may play a strategic role in the language information network system to be conceptualized by the LINCS project. The ultimate configuration of centers or "nodes" within this network will depend on numerous factors which are now being studied by the project.

Before passing to a discussion of some of the developmental activities of the LINCS project, it might be well to summarize briefly the major problems besetting the formal channels of information transmission already in existence, and to which the designers of any large-scale system for the future must address themselves.

(1) The multiplicity of languages used makes a good deal of the literature relatively inaccessible to at least some users, and difficult to monitor and process in secondary and tertiary services.

(2) There is, in the latter, a great deal of waste through duplicated effort. This is particularly serious in a field in which monetary resources tend to be much more limited than in the "hard" sciences.

(3) The "peripheral" literature is very widely scattered and hence difficult to locate.

(4) Not previously mentioned, but constituting a serious problem, is a lack of effective basic tools -- dictionaries, classification schemes, thesauri -- needed to impose structure on the literature.

(5) There is no central abstracting-indexing service. A high degree of idiosyncrasy is found in coverage policies of secondary services.

(6) Coordination of effort and cooperation are minimal.

Some of these problems are primarily technical and must be overcome through more sophisticated techniques; others could at least be improved through greater organization and coordination.

4. Developmental Activities: The LINCS Project

The LINCS project is by no means the only attempt to bring the benefits of modern technology to bear on the problems of information transfer in the language sciences. Certainly the information centers I have described, and the many others I have referred to, share with us this aim to one extent or another. What distinguishes our project from the others is its scope: to the best of our knowledge, it is the only program aimed at serving the entire language-sciences community through control of all literature germane to the interests of any part of that community. The other centers to date have often tended to be mission-oriented; LINCS will be discipline-oriented and will define its discipline as comprehensively as possible.

The LINCS project, which is supported by the National Science Foundation, began with a survey-and-analysis stage (1967-68). This was followed by a preliminary system-design stage (1968-69) and the current advanced system-design stage, which will be completed in July 1971. Thereafter, a system acquisition or implementation phase of four to five years is envisioned.

The project is placing a special emphasis on the development and demonstration of its representativeness (agency or mandate), responsibility, and readiness -- three fundamental requirements.

On the face of things, evidence of a mandate from a heterogeneous community should be very difficult to demonstrate. To an extent, this has been true; the absence of any "umbrella" organization has increased the complexity of the project's relations with its constituency. As I have already pointed out, however, one of the reasons for the establishment of CAL was the recognized need for varied information services; the actual provision of such services remains one of its most important functions. Its mandate to pursue the objectives embodied in the LINCS project has arisen through CAL's continuing relationships with the major professional organizations in the language sciences, which have voiced interest in, and support of, our undertaking. We are, of course, working to develop further this interest and support, to the extent that effort on this constitutes one of the major subtasks of the current stage.

The LINCS project has, throughout its two-year history, had a two-pronged approach: we have simultaneously pursued the definition of the main goals of an information system and a study of the techniques required to attain the emerging objectives.

In the first area, we began with an introductory examination of planning approaches. We made preliminary samplings of characteristics of the user community and of existing information channels. We studied the problems of conceptual alternatives, of potential interfaces between a LINCS and other information systems, and of various techniques for the planning and management of a LINCS.

These efforts carried forward into the program's second stage, in which we began to work toward a more explicit formulation of the service objectives of a LINCS. In this stage, we have attempted a statistical definition of the potential user community, and have begun to compile data on the behavior of various segments of that community. We have developed sample data on journals and citations for use in a more thorough investigation of formal channels of information transfer. And we have conducted a preliminary examination of some economic and technical requirements of system alternatives.

I have highlighted a few of the findings of these activities for you today. When complete (as most of them will be very shortly), they will lead us into the next step in this phase of our work. In it, we will concentrate on: (1) an exhaustive description of the current communication system; (2) definition of the system concept and preparation of an implementation plan; and (3) the development of various program management capabilities, including a management information system and, as I have mentioned, development of the professional community's advisory functions.

In addition, our project began with a general survey of high priority components for a LINCS. We looked at various indexing systems and terminologies, and acquainted ourselves with the general problems involved in system automation. In the project's second stage, we collected as many relevant thesauri as possible, continued our investigation of alternative indexing systems, and began work on a preliminary LINCS thesaurus, following the Committee on Scientific and Technical Information (COSATI) Guidelines[7]. The thesaurus will have two components: the scientific terminology and a list of language names. Our listing of language names is the most complete yet devised, and it is anticipated that the Library of Congress might adopt it; it contains about 18,000 entries. Work on the thesaurus and development of a retrieval capability will continue in the LINCS project's third stage.

---

[7] Guidelines for the Development of Information Retrieval Thesauri, prepared by Sub-Panel on Classification and Indexing, Panel on Operational Techniques and Systems, Committee on Scientific and Technical Information (COSATI) (Washington, D.C., 1967).

In the area of system automation, our initial study of input, storage, transmission, format, and typography requirements led to a survey of file-management techniques in general and of some particular operational systems. We have undertaken a major study of typographic and stylistic characteristics of documents in the language sciences. Third, we have begun to study problems associated with compatibility and standardization. These will continue in Stage Three.

We will, moreover, acquire some "real-world" experience, including marketing details, through the operation of several experimental publication systems. The exact nature of these will not be determined until we have analyzed more data on user needs and interests, but we do intend to cover the entire range of primary, secondary, and review publications. Most probably, we will test several alternatives in each category. In addition to giving us experience in actual processing tasks, we should obtain feedback useful in the specification of the system concept, and learn something of the relative value of different marketing techniques. As a preliminary to actual marketing studies, we are at present constructing lists of potential audiences, on the one hand, and possible products and services on the other. Each of these will be specified in increasing detail, and a hierarchy of priorities will be established. When this has been accomplished, mock-ups of products will be tested on selected sample audiences.

5. Some Problems of System Design

Our efforts to date have already posed a number of problems that must be overcome if our system design effort is to be successful. I would like to conclude this talk by describing to you a couple of thses problems.

One of the most formidable has emerged from our examination of data elements in bibliographic records and bibliographic and typographic conventions. It involves the size of the character set required for any kind of publication in the language sciences, and special graphic features found in primary publications. The problem arises from several sources. First, as I have already mentioned, the language science literature occurs in an unusually large number of languages (the latest issue of the international Linguistic Bibliography cited documents in 50 languages). Second, the number of languages that may be viewed as subject matter increases the total number of languages still further: if we add to the 50 languages cited in the LB the languages embedded in the citations, the total figure rises to

about 90. Many of these languages employ diacritical variants of the Roman alphabet; a number use other alphabets entirely. Assuming some kind of photocomposition in our future publications processes, this means we are faced with a serious impediment. Restricting ourselves to bibliographic publications, we estimate a minimum set (ignoring differences of type style) of about 1,000 unique characters.

In dealing with the primary literature, our requirement will be considerably higher. In addition to a more extensive use of diacritics, the primary literature is distinguished by a much more frequent occurrence of special symbologies used in the phonetic and phonemic transcription of languages and dialects. Since these symbologies are rarely used in titles, we could probably get by, in our bibliographies, with about 100 special characters devoted to this purpose. In handling primary literature, this number would have to be much larger; just how much so, it may not be possible to determine with precision, since the use of these characters depends on the precision of sound representation that the linguist wishes to achieve. A linguist, for the narrowest transcription, uses about 200-250 symbols.

An added complication in dealing with the primary literature will be the number of special graphic features required. These are used to display relationships among sounds, syntactic elements, dialects, languages, and language groups.

No ready solution has presented itself. With sophisticated electronic character generating equipment, the character set is theoretically unlimited, but consumption of time and money required for the creation of special characters represents a very real practical problem. Moreover, this would still leave difficulties in inputting and generating output for anything except hard copy.

Standardization, transliteration, and other more or less arbitrary means of reducing the size of the character set may represent a partial solution. They can only be carried to a certain point, however, without sacrificing accuracy; determination of where this point lies would have to be the subject of very careful study. Moreover, the promulgation of standards in a field as cosmopolitan as ours would be difficult even if the community were very effectively organized. To be acceptable, such standards must avoid the impression of strong association with a single constituent; for an area like transliteration, this become highly problematic. Imposition of such standards would bring with it problems in conversion, for either LINCS, its users, or both. Moreover, at present there has not been a great deal accomplished in the way of establishing international standards that would be helpful to us in this area.

Our investigation of indexing and retrieval has established several clear constraints. Users of the system are expected to represent highly interdisciplinary interests, needs, and points of view; they are expected to be native speakers of a number of different languages. In addition to having to take account of vocabulary problems presented by these factors, there is the additional point to remember that we are dealing with a relatively "soft" literature, with low standardization of terminology.

For these and numerous other reasons, the thesaurus format was chosen as the most advantageous type of indexing language for LINCS. As I have already told you, we have collected samples of thesauri relevant to a LINCS and begun some experimentation. We have been guided by the USA Standard: Basic Criteria for Indexes[8] and the COSATI Guidelines for the Development of Information Retrieval Thesauri. We are continuing our examination of a variety of techniques and approaches, including use of macro- and microthesauri. At the same time, we are contemplating collaboration on the revision of Class 8 (linguistics) of the Universal Decimal Classification (UDC), which we feel might be used at some of the international interfaces of LINCS.

Let me close by adumbrating the kind of system we have been led to visualize. I must qualify this by saying that this formulation is intentionally imprecise: our own conception is only partially formed, and will gain clarity only through further analysis of the data already collected, continued research, and actual experimentation.

We envision the system as being integrated in two major dimensions: a vertical, axial array of functional components (various kinds of processing), and, radiating out from this, a network of actual services. Horizontal integration will result in networks of the various functional components; vertical integration will provide free flow among functionally different components. Such integration will be provided through various linkages, about which data have been collected, but not yet analyzed. This network arrangement will most probably be relatively loose: many of the individual nodes may also participate in other information systems, as may the system as a whole. The integration of the functional components will be the work of the central system authority, which will serve as a switching facility for the entire network. Our evaluation of alternative arrangements of components, and implementation of the system concept, will be based on the degree to which they promote and maintain integration on these two parameters.

---

8 USA Standard: Basic Criteria for Indexes, USAS Z39.4-1968, revision of Z39.4-1959 (New York: United States of America Standards Institute, 1969).

Implicit in this requirement is a consideration of economic realism. We expect the final operating network to be capable of self-support (the principle of synergism, which ought to have strong positive implications in an endeavor of this sort, can be seen in our requirement of vertical integration); this requirement of self-support will be one of the ultimate measures of effectiveness of the design.

We expect the work of system design to be completed in about one year, and the system to be fully operational by 1975.